Stereo Seam Carving A Geometrically Consistent Approach

Tali Basha, *Student Member, IEEE,* Yael Moses, *Member, IEEE,* and Shai Avidan, *Member, IEEE*

Abstract—Image retargeting algorithms attempt to adapt the image content to the screen without distorting the important objects in the scene. Existing methods address retargeting of a single image. In this paper we propose a novel method for retargeting a pair of stereo images. Naively retargeting each image independently will distort the geometric structure and hence will impair the perception of the 3D structure of the scene. We show how to extend a single image seam carving to work on a pair of images. Our method minimizes the visual distortion in each of the images as well as the depth distortion. A key property of the proposed method is that it takes into account the visibility relations between pixels in the image pair (occluded and occluding pixels). As a result, our method guarantees, as we formally prove, that the retargeted pair is geometrically consistent with a feasible 3D scene, similar to the original one. Hence, the retargeted stereo pair can be viewed on a stereoscopic display or further processed by any computer vision algorithm. We demonstrate our method on a number of challenging indoor and outdoor stereo images.

Index Terms—Stereo, Retargeting, Geometric Consistency.

---

1 INTRODUCTION

Digital images are displayed on a variety of digital devices, each of which might require a different aspect ratio. The core idea of image retargeting algorithms is to adapt the image content to the screen without distorting the important objects in the scene. The rapid pace of technology makes it possible to view 3D content on a large range of devices, from cellphones to large TV screens. In addition, stereophotography is becoming increasingly popular, with a large number of stereo images appearing online. As a result, image retargeting algorithms need to be adapted to work on stereo image pairs.

We propose a novel method for retargeting stereo image pairs. The input to our method is assumed to be a rectified stereo image pair and a disparity map. The input disparity map may be computed from the pair of images by an available stereo algorithm, or be given by any other algorithm or device. The 3D information provides valuable cues for retargeting, as previously demonstrated by retargeting algorithms for a single image [2]. Indeed, stereo image retargeting can also benefit from the 3D information provided by the other image. However, since the 3D information must be maintained in the retargeted pair, maintaining the 3D information poses new challenges.

Our method retargets the input pair in the horizontal domain while minimizing the distortion of each image as well as the distortion in depth. A key property of our method is that the retargeted stereo pair has a feasible 3D interpretation that is similar to the original one. Thanks to this geometric consistency, our retargeted pair can be viewed on a stereoscopic

---

- T. Basha & S. Avidan are with the Department of Electrical Engineering, Tel-Aviv University, Tel-Aviv, 69978, Israel.
  E-mail: talib@eng.tau.ac.il
- Y. Moses is with the Interdisciplinary Center, Herzliya 46150, Israel.

display or processed by any computer vision algorithm that makes use of a stereo pair (e.g., cosegmentation or tracking).

1.1 The General Idea

We generalize the single image seam carving algorithm [3], [4] to work on a rectified stereo pair. Instead of removing a seam from a single image, our algorithm iteratively removes a pair of seams from the stereo image pair.

A naive extension of the single image seam carving algorithm is to independently apply it to each of the images (see the blue box in Figure 1**a**). It disregards the geometry and as a result, damages the 3D structure of the scene, for example by removing a pixel from one image while keeping its corresponding pixel in the other one. To overcome this problem, a joint retargeting of both images must be considered. In particular, the selection of seams in both images should be coupled. A straightforward approach for stereo retargeting is to compute the seam in one of the images, say the left one, and then map it to the right image via the disparity map. This is clearly sub-optimal as it does not utilize the information available in the right image or the depth map. Figure 2 demonstrates the results using this approach (for more details see Section 6).

In fact, the problems run even deeper, mainly due to occlusions; there is no guarantee that seam pixels in the left image have matching pixels in the right. And the change in 3D shape must be carefully considered to avoid an inconsistent change in the visibility relation of the scene points. In particular, pixels that are visible only in one of the views should not be revealed. Thus, the new problem of stereo retargeting creates new challenges.

The proposed method overcomes the challenges of stereo retargeting by generalizing seam carving to simultaneously carve a pair of seams in both images, while minimizing distortion in appearance and depth. Seam selection and disparity map
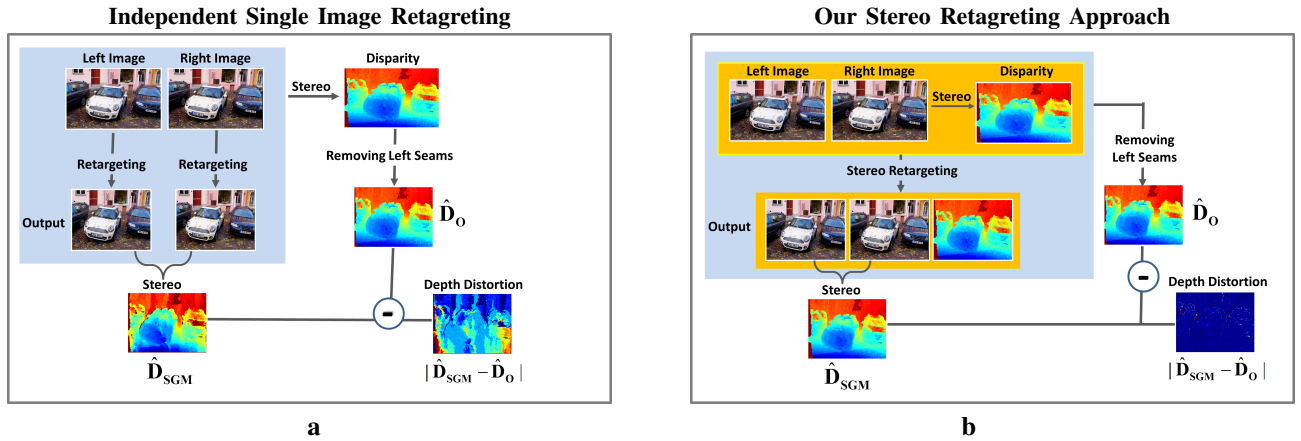
Fig. 1. **Geometric Evaluation. (a):** The results of applying single image SC [4] on each of the input images.**(b):** The results of applying our stereo retargeting algorithm. On both left and right sides: $\widehat{D}_{SGM}$ is computed by applying SGM [5] on the retargeted pair; the original disparity values of the remaining pixels are stored in $\widehat{D}_o$. Finally, the depth distortion is measured by $|\widehat{D}_{SGM} - \widehat{D}_o|$. The color code is blue for low values and red for high ones; red indicates a difference of at least six pixels.

modification are subject to geometric constraints that take into account the visibility relations between pixels in the images (occluded and occluding pixels). These geometric constraints *guarantee* consistency of the target images with a feasible 3D scene, as formally proven in Section 5 and empirically demonstrated in Section 6.

2 BACKGROUND

Image and video retargeting algorithms have been extensively investigated in recent years. These algorithms attempt to change the aspect ratio of an image or a video in a way that does not distort the proportions of the important objects in the image. The various algorithms differ in how they determine the importance of different pixels in the image and in how they use this information. Two main classes of algorithms have emerged. Discrete methods for single image retargeting, such as seam carving [3] or shift map [6], remove and shift pixels in the image. Continuous methods [7], [8] warp a quad mesh based on image content. An excellent overview and comprehensive study of the topic is given in [9].

Here, we extend the seam carving algorithm to work on stereo. The algorithm was first introduced in [3] and was extended in [4]. The seam carving algorithm works by iteratively computing a seam with minimal visual distortion in the image and removing it. A seam is defined to be a connected path in the image, yet this is not a necessary assumption and Grundmann *et al.* [10] recently showed that piece-wise connected paths are more flexible for video retargeting. There the goal is to retarget frames sequentially and rely on piece-wise connected seams to better fit the retargeting to previous frames. However, they do not consider stereo data.

Most work reported thus far in the literature has focused on retargeting a single image or video. However, the rise of 3D content makes it necessary to extend image retargeting algorithms to work with 3D content. Lang *et al.* [11] adjust the disparity map according to various stylistic considerations. They do not consider the problem of stereo image retargeting,

nor do they discuss the the geometric consistency of their method.

Mansfield *et al.* [2] assume the input is a single image and a relative depth map (provided by the user) and the output is a single image. They extend seam carving to *scene carving* and show that scene carving is indeed scene consistent, can introduce occlusions, and can also handle pixel reappearance, say when one layer moves behind another layer and reappears on the other side. There are a number of important distinctions between our work and that of [2]. First, we assume that the input is a pair of stereo images. They, on the other hand, assume the input to be an image with a depth map. As a result, they cannot produce a retargeted stereo pair without resorting to image synthesis techniques to fill in gaps, as they have to deal with the occlusions caused by the different points of view of the stereo image pair. Second, we assume a per-pixel stereo map, as opposed to representing the scene as a collection of well-defined fronto-parallel planes.

Chang *et al.* [12] proposed a content-aware display adaptation method that simultaneously resizes a stereoscopic image to the target resolution and adapts its depth to the comfort zone of the display while preserving the perceived shapes of prominent objects. This is done by detecting and matching a sparse set of feature points that are then used to define a warping field according to the target display parameters.

In the scene-warping method by Lee *et al.* [13], the layer-based approach [2] and the warping-based approach [12] are combined. Each of the input stereo images is decomposed into multiple layers according to color and depth information. Each layer is then warped by its own mesh deformation, and the warped layers are composited together to form the resized images. Both methods [12], [13] do not discuss the geometric consistency of their method, nor how they deal with occlusions between the left and right views.

Utsugi *et al.* [14] have also considered the extension of seam carving to stereo images. However, here too, the primary goal of preserving the geometric consistency of the output image
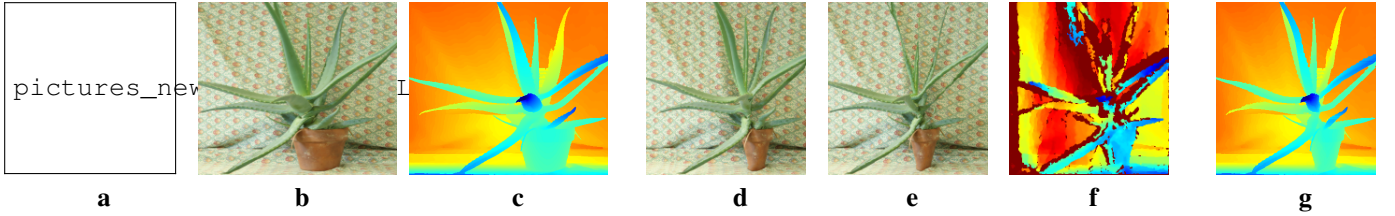
Fig. 2. **The straightforward approach for seam coupling. (a-b)**: The pair of input images, and **(c)**, their ground truth disparity map. **(d):** The single image retargeted left image using [4], and **(e)** the retargeted right image computed by carving corresponding seams. **(f):** The disparity map computed by applying the SGM stereo algorithm on the pair **(d-e)**. **(g):** The updated disparity map computed by our method.

pair is neither defined nor discussed in their work.

Recently, Birklbauer & Bimber [15] proposed a method for light-field retargeting that preserves the angular consistency. Their algorithm converts a stack of images into a light field representation in which seam carving is performed. The retargeted light field is then mapped back to retargeted individual images. They avoid the need to recover explicit 3D information, however it is not clear how they deal with occlusions.

Finally, our work is part of a recent surge in stereo image editing algorithms, where the goal is to extend image editing tools to work directly on a pair of stereo images. This includes inpainting of stereo image pairs [16], stereoscopic 3D cut-and-paste [17], and viewer-centric editor for 3D movies [18].

3 THE METHOD

The input to our method is a pair of $m \times n$ rectified stereo images, $\{I_L, I_R\}$, and a disparity map, D, where the disparity map can be computed by any stereo algorithm (we use [5]). Without loss of generality, we consider the disparity with respect to the left image, which is taken to be the reference image. The output of our algorithm is a pair of retargeted images, $\{\widehat{I}_L, \widehat{I}_R\}$ and an updated disparity map, $\widehat{D}$.

The primary goal of our method is to obtain retargeted images that are geometrically consistent with a feasible 3D scene. Namely, it is possible to define correspondence between pixels that is consistent with the epipolar geometry as well as with visibility relations between 3D points in the retargeted pair. To obtain this goal it is sufficient to impose the following constraints:

C$_1$: *Corresponding pixels in the original images are either both removed or remain corresponding in the output images.*

C$_2$: *3D points that are visible in the reference view but occluded in the other are not revealed.*

In Section 5 we formally prove that these constraints are satisfied by our method.

3.1 Seam Coupling

The geometric coupling of the two seams, $S_L = \{s_L^i\}_{i=1}^m$ and $S_R = \{s_R^i\}_{i=1}^m$, is obtained by using the correspondence defined by D. Formally, each of the seam's pixels in the left image at row i, $s_L^i = (i, j_L(i)) \in S_L$, is matched to a seam pixel in the right image, $s_R^i = (i, j_R(i)) \in S_R$, as follows:

$$s_R^i = (i, j_R(i)) = (i, j_L(i) + D(s_L^i)), \quad (1)$$

where $j_L, j_R : [m] \to [n]$, and $[m] = [1, ..., m]$. The estimated disparity map, $D : [n] \times [m] \to \mathbb{Z} \cup \perp$, maps pixels of I_L to their corresponding pixels in I_R, if the correspondence is known, and to $\perp$ otherwise. Note that the seams contain only pixels for which the disparity is defined.

Note that a continuous seam in the left image generally corresponds to a piecewise continuous seam in the right image since the seam may cross depth discontinuities. Therefore, we drop the assumption that a seam (in either I_L or I_R) is continuous and consider piecewise seams from now on, which we refer to as *generalized seams* (see Figure 3).

3.2 The Energy Function

The energy function of the stereo seam carving method consists of an intensity term and a 3D geometry term. Removing a seam's pixel from each image in the stereo pair has the local effect of creating new adjacent pixels in the target image. The resulting gradients in the retargeted left and right images depend on the seam pixel in the previous row, denoted by $j_L^\pm$ and $j_R^\pm$, respectively. Since the left and right image seams are coupled, $j_R^\pm$, is uniquely defined by $j_L^\pm$ and the disparity map, D. Therefore, we define the energy function (w.r.t. the left image) in accordance with the seam pixel in the previous row, $j^\pm$ (which is short for $j_L^\pm$). That is,

$$E_{total}(i, j, j^\pm) = E_{intensity}(i, j, j^\pm) + \alpha E_{3D}(i, j, j^\pm), \quad (2)$$

where α controls the relative impact of each of the terms. Since we use generalized seams, $j^\pm \in [m]$ can be any pixel in row $i - 1$ (unlike the continuous case in which $j^\pm \in \{j - 1, j, j + 1\}$).

3.2.1 Appearance Energy

We generalize the *forward energy* criterion from [4], whose goal is to minimize the resulting distortion in the retargeted image caused by the intensity differences between new adjacent pixels. The appearance distortion $E_{intensity}(i, j, j^\pm)$ is taken to be the sum of the energy terms, E_L and E_R, for removing a pair of coupled pixels from the left and right images. That is,

$$E_{intensity}(i, j, j^\pm) = E_L(i, j, j^\pm) + E_R(i, j_R, j_R^\pm), \quad (3)$$

where the coupling of the left and right seams is captured via the disparity map as defined in Section 3.1.

The energy of removing a specific pixel, (i, j) from image I, left or right, is given by:

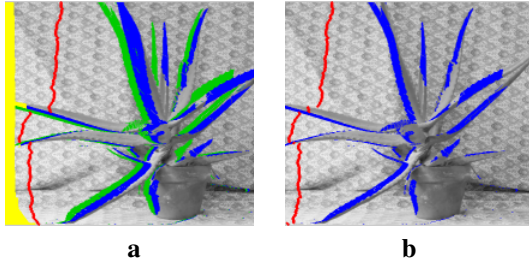$$E(i, j, j^\pm) = E^v(i, j, j^\pm) + E^h(i, j), \quad (4)$$

Fig. 3. **(a)**: The left image masked with the computed occluded pixels in green, and occluding pixels in blue; out-of-field-of-view pixels are colored in yellow. **(b)**: The right image masked with the corresponding occluding pixels in blue. In this example both left and right seams (in red) are discontinuous.

where E^h and E^v are the forward energy terms due to the new gradients in the horizontal and vertical directions, respectively. In particular, E^h is given by:

$$E^h(i,j) = |I(i, j+1) - I(i, j-1)|. \qquad (5)$$

In the vertical direction, the new gradients depend on the position of the seam in row $i-1$, $j^{\pm}$. Accordingly, the vertical forward energy is given by:

$$E^v(i, j, j^{\pm}) = \begin{cases} V_1 & j^{\pm} < j \\ 0 & j^{\pm} = j \\ V_2 & j^{\pm} > j \end{cases} \qquad (6)$$

where

$$\begin{aligned} V_1 &= \sum_{k=j^{\pm}+1}^{j} |I(i-1, k) - I(i, k-1))| \\ V_2 &= \sum_{k=j+1}^{j^{\pm}} |I(i-1, k-1) - I(i, k))|. \end{aligned} \qquad (7)$$

3.2.2 Depth Energy

The computed depth map provides valuable cues for seam selection, and a 3D forward energy term, E_D, is used to minimize the disparity distortion. It is defined similarly to the forward energy of the intensity values, by replacing the intensity function, I, with the disparity map D in Eq. 4 and Eq. 7. In practice, in order to compensate for the differences in range between the intensity and the disparity values, we normalize both I and D in the range of zero to one.

In addition, the object's distance from the camera often correlates with its saliency. Hence, we increase the energy of pixels that are the projections of nearby 3D points. Moreover, our method is strongly based on the disparity map, which is computed by a stereo algorithm that is regarded as a black box. Errors in the estimated map may result in incorrect coupling of seam pixels. We prefer removing pixels for which we have high confidence of disparity values, measured by the difference in the intensities of corresponding pixels. That is,

$$G(i,j) = |I_L(i,j) - I_R(i, j + D(i,j))|. \qquad (8)$$

The total forward 3D energy is a weighted sum between three components:

$$E_{3D}(i, j, j^{\pm}) = E_D(i, j, j^{\pm}) + \beta |D_n(i,j)| + \gamma G(i,j), \qquad (9)$$

where D_n is the normalized disparity map.

3.3 Maintaining Pixel Visibility

An *occluded* pixel in the reference image is defined as the projection of a 3D point that is not visible in the right view due to another 3D point that occludes it (red in Figure 4**a**). Occluded pixels do not have corresponding pixels in the right image; our method does not remove them from the image.

Furthermore, in order to satisfy the geometric constraint, $\mathbf{C}_2$, occluded pixels must not be revealed. Otherwise, no coherent 3D interpretation can justify the visibility of the revealed pixel only in one image and not in the other. To this end, we ensure that occluded pixels in the original right image remain occluded in the retargeted right image, by avoiding removing pixels that may reveal them, namely *occluding* pixels. An occluding pixel is defined to be the projection of a visible 3D point in both views that accounts for the occlusion of one or more 3D points in the right view (see green in Figure 4**a**). Our choice of removing only pixels that are neither occluded nor occluding, guarantees that the original visibility relation (i.e., occluded-occluding pairs) is preserved. See Section 5 for the proof.

The set of occluding and occluded pixels is computed once from the input disparity map, D and represented by a binary map, $O(i,j)$ where $O(i,j) = 1$ if pixel (i,j) is an occluded or occluding pixel. This map is computed using a simplified Z-buffer approach. Namely, if two or more pixels in the left image are mapped to the same pixel in the right image, the pixel with the largest disparity value is the occluder while the rest are occluded.

In the examples we considered, the number of occluded and occluding pixels is typically 20%. An example of both the occluded and occluding maps is given in Figure 3.

3.4 Stereo Seam Selection and Carving

The energy term defined in Eq. 2 is now accumulated in a cost matrix M to select a pair of seams using dynamic programming. The seams are coupled as defined in Section 3.1. We
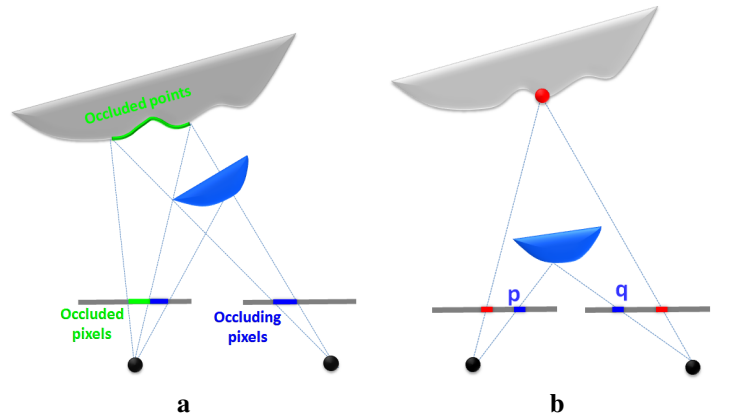

Fig. 4. **(a):** Occluded pixels, green, have no corresponding pixels in the right image. The occluding pixels, blue, are visible in both views. **(b):** The ordering constraint does not hold: removing the red point causes the point *p* to shift left while the point *q* remains in its original location.
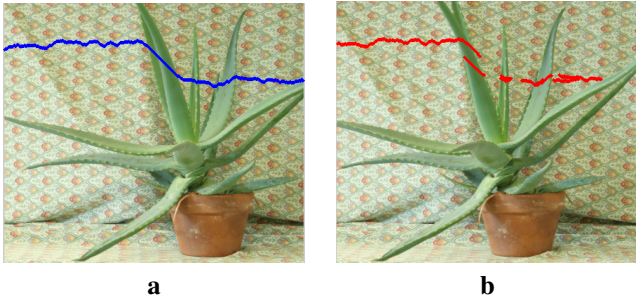
Fig. 5. **Horizontal Seams:** A horizontal seam in the left image (**a**) does not specify a seam the right (**b**).
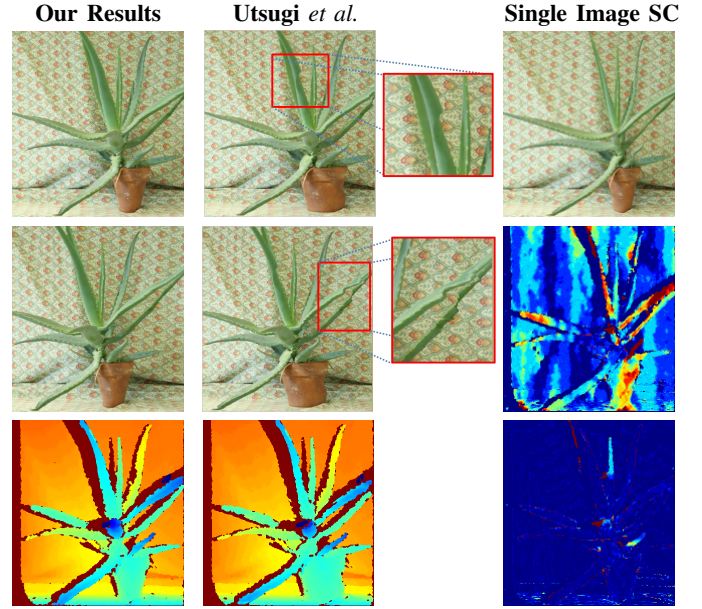

Fig. 7. **Aloe Dataset.** In the first column (top to bottom), our results for the left and right images and the disparity map. The input stereo pair is shown in Figure 2(**a-b**). In the second column, the results of the method of Utsugi *et al.*[14]. In the third column, the results of applying single image SC [4] to the left input image (top); see caption of Figure 6. The depth distortion scores: single image SC, B=47%; stereo pair SC, B=2.9%.

set $M(i,j) = \infty$ for pixels that do not satisfy the visibility constraints, namely if $O(i,j) = 1$ (see Sec 3.3).

By default, we prefer continuous seams (where $j^\pm \in j-1, j, j+1$), which affect fewer pixels than discontinuous seams (see Eq. 7). However, if a continuous path is blocked at pixel (i,j) by occluded/occluding pixels, we allow discontinuous seams. Formally, we consider two cases, according to whether it is necessary to switch at the pixel (i,j) from a continuous to discontinuous seam:

$$
M(i,j) = \begin{cases} \min_{j^\pm \in \{j-1,j,j+1\}} E_{total}(i,j,j^\pm); & T(i,j) = 0 \\ \min_{j^\pm \in [m]} E_{total}(i,j,j^\pm); & T(i,j) = 1, \end{cases}
$$
(10)

where, T is the binary map of size $n \times m$. $T(i,j)$ indicates whether a continuous path is blocked in row $i-1$ by occlud-
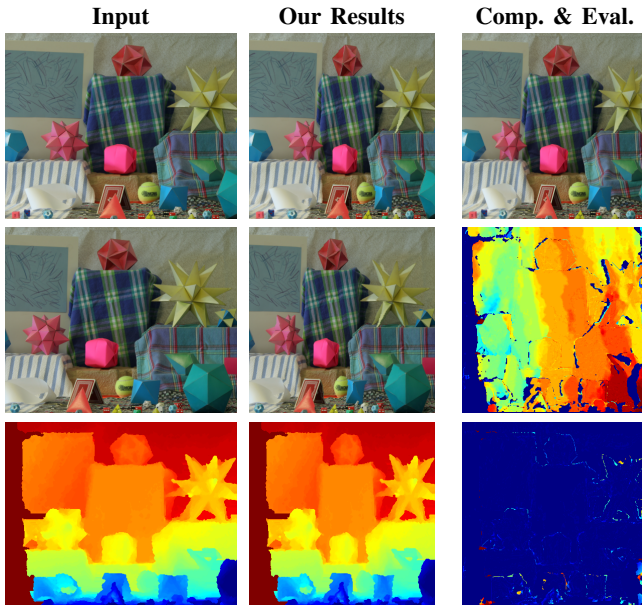

Fig. 6. **Moebius Dataset.** In the first column (top to bottom), the input left and right images and the input disparity map. In the second column, our results, with respect to the first column. The third column shows the results of applying single image SC [4] to the left input image (top); the distortion in depth caused by independent single image retargeting (middle); the distortion in depth caused by our stereo retargeting method. Depth distortion scores: single image SC, B=85%; stereo pair SC, B=3.2%.
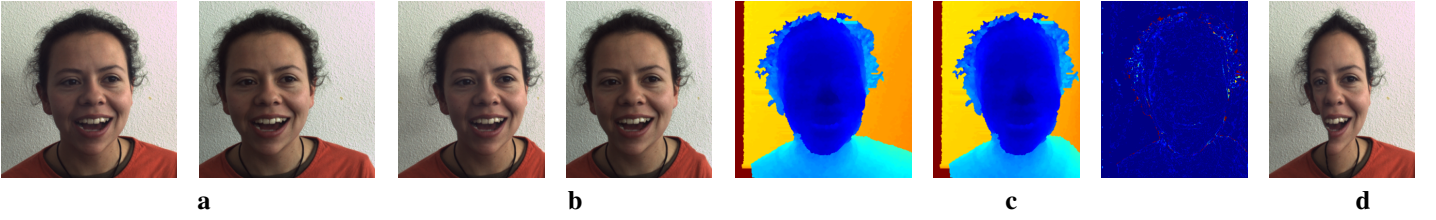
ing/occluded pixels. That is, $T(i,j) = 1$ if $O(i-1,j^\pm) = 1$ for $j^\pm \in \{j-1, j, j+1\}$. Note that piece-wise connected seams were successfully used for video retargeting [10], where the goal is to preserve the moving regions.

As in [4], removing a seam pixel from a row results in shifting pixels in that row. Specifically, all pixels to the right of the removed pixels are shifted left by one pixel. The remaining pixels are unchanged. Formally, the shifting function $f_L(i,j) : [m] \times [n] \to [m] \times [n-1]$ maps the ith input row to the ith output row. Let $s_L^i = (i, j_L(i))$ be the pixel to be removed from the left image. Then, the shifting mapping is defined by:

$$
f_L(i,j) = \begin{cases} j & \text{if } j < j_L(i) \\ j-1 & \text{if } j > j_L(i) \\ \bot & \text{if } j = j_L(i) \end{cases}
$$
(11)

Likewise, $f_R(i,j)$ is the corresponding mapping function in the right image, where $j_L(i)$ is replaced by $j_R(i)$ (as defined in Eq. 1).

After carving a seam, the new disparity map, $\widehat{D}$, is obtained by removing the left seam S_L from the previous D and updating the disparity values of the remaining pixels. In particular, the updated disparity value, $\widehat{D}$, of a pixel (i,j) is given by:

$$
\widehat{D}(i, f_L(i,j)) = f_R(i, j + D(i,j)) - f_L(i,j).
$$
(12)

3.4.1 Geometric Interpretation

We next describe the geometric interpretation of the carving. From Eq. 11 it follows that each pixel may either be shifted one pixel to the left or remain in its original location. If a

Fig. 8. **Diana Dataset. (a):** The input pair of images. **(b):** The retargeted pair of images produced by our method. **(c):** The input disparity map, our result, and the evaluated depth distortion. **(d):** The result of single image SC on the left image.

pair of corresponding pixels remains in its original location, the associated 3D point remains the same as in the original scene. When pixels are shifted, the position of the associated 3D points change. If the two pixels in a corresponding pair are both shifted left, the original depth is preserved, namely $\widehat{D}(i,j) = D(i,j)$ (see Eq. 12). The associated 3D point changes its location accordingly by a left translation, parallel to the image plane. Most pixels will either remain in their original location or be shifted together. However, when the ordering constraint does not hold (see Figure 4**b**), a pixel may be shifted in one of the images, while its corresponding pixel remains in its original location. In this case, the disparity is changed by one pixel, which corresponds to a small change in depth.

3.5 Stereo Image Pair Enlarging

So far we have shown how to reduce the width of the input stereo pair, but our method can also be applied to enlarge the width. This is done, similarly to the single image seam-carving algorithm [4], by first selecting the optimal pairs of seams for removal, and duplicating them in the pair of images. In addition, we update the disparity map by duplicating the left image seams, and updating the disparity values when necessary (i.e., corresponding pixels in the left and right images are not on the same side of the seam in both images.)

4 HORIZONTAL SEAMS

Existing single image retargeting methods can be directly applied to change the height of the image as well since the vertical and horizontal directions are symmetric. However, when a stereo pair is considered this is no longer the case. Preserving the geometric consistency of a pair of images is possible only if restrictive assumptions on the disparity map are imposed, as we describe below.

To preserve the 3D interpretation of the scene, it is essential to couple the seams of the pair of images (see Section 3.1). As in the vertical case, a seam in the left image is mapped to a set of corresponding pixels in the right image, determined by the disparity map. However, in the horizontal case, this set of pixels generally does not specify a horizontal seam in the right image. That is, there is no guarantee that the set of corresponding pixels in the right image consists of a single pixel at each column (see example in Figure 5). An exceptional case is a constant disparity along the seam, where the right image seam is simply a shift of the left one. However, even under this restriction (which is not valid in practice),

the disparity must be further constrained for retargeting while maintaining the geometric consistency of the stereo pair. In particular, pixels above the seam remain at the same location, while pixels below the seam are shifted up in one pixel. Thus, the epipolar lines of the pixels below the seam are shifted up as well. It follows that in order to preserve the epipolar geometry, corresponding pixels in the left and right image must be on the same side of the seam (below or above) in both images.

We conclude that constraints on both the seam disparity and on the disparity of pixels in the region bounded by the rows that participate in the seam are unlikely to be satisfied for realistic scenes. Hence, horizontal seams cannot be used for retargeting while keeping the geometric consistency of a stereo pair. It is worth noting that the image height can be resized by expanding the image width, followed by uniform scaling. Clearly, this is not optimal but allows the aspect ratio to be changed as desired.

5 GEOMETRIC CONSISTENCY

We prove here that our algorithm preserves the geometric consistency of the input pair. Clearly, the epipolar geometry is preserved (as well as the rectification) since all pixels in both images are either left-shifted or remain in their original location. We next show that our method satisfies the constraints C_1 and C_2 (see Section 3).

Constraint C_1, maintaining the original input correspondence, is directly satisfied by our method since the disparity map is used to couple the seams (see Section 3.1). To show that constraint C_2 is satisfied, we prove that the operation of removing a pair of seams pixels and the following update of the disparity map guarantee that the original visibility of 3D points is preserved. To this end we formally define the *occluded* and *occluding* pixels.

Definition: Occluded and Occluding Pixels

Let (i, j_b) and (i, j_f) be two pixels in the left image. The pixel (i, j_f) occludes (i, j_b) *iff* $j_b < j_f$ and the two pixels are mapped to the same pixel in the right image. That is,

$$j_f + D(i, j_f) = j_b + D(i, j_b). \tag{13}$$

It follows that a pixel (i, j) is *not* an occluding\occluded *iff*

$$j + D(i, j) \neq j' + D(i, j') \quad \forall j' \neq j. \tag{14}$$

Lemma #1:

The operation of removing a seam point, $p_L = (i, j_L)$, preserves the ordering between the remaining pixels in the row
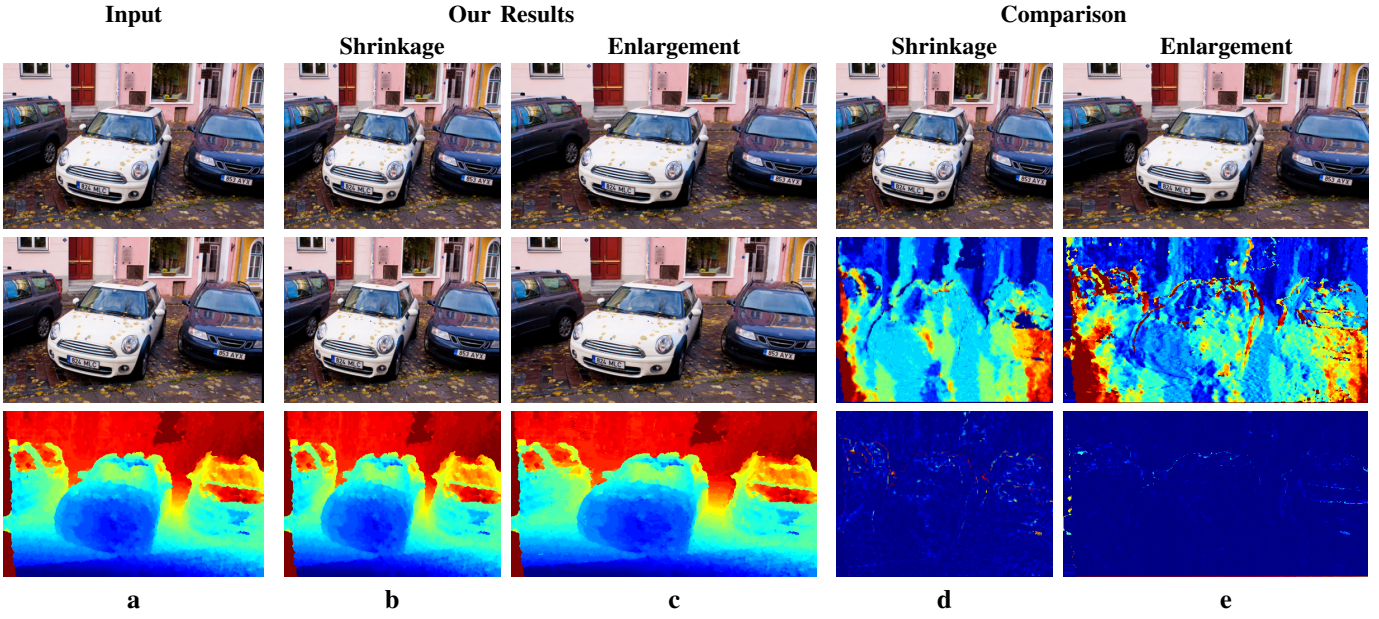
Fig. 9. **Car Dataset.** In column (**a**) (top to bottom), the input left and right images and the input disparity map. In columns (**b**) and (**c**), our results for reducing and increasing the width by 17%. In columns (**d**),(**e**), the results of applying single image SC to the left input image (top); the distortion in depth caused by independent single image retargeting (middle); the distortion in depth caused by our stereo retargeting method. Depth distortion scores: Shrinkage: single image SC, B=70%; stereo pair SC, B=1.3%; Expansion: single image SC, B=69.3%; stereo pair SC, B=1.9%.

i. Formally, given two pixels, $p_1 = (i, j_1)$ and $p_2 = (i, j_2)$: $j_1 < j_2 \Leftrightarrow f_L(i, j_1) < f_L(i, j_2)$, where f is defined in Eq. 11.

It follows directly from this Lemma that: $j_1 = j_2 \Leftrightarrow f_L(i, j_1) = f_L(i, j_2)$.

Proof: If both p_1 and p_2 are on the same side of the seam, then by Eq. 11 the order is preserved. Therefore, the only case to consider is when the seam pixel, (i, j_L), is in between the two pixels: without loss of generality, $j_1 < j_L < j_2$. In this case, $f_L(i, j_1) = j_1$ and $f_L(i, j_2) = j_2 - 1$. Since this scenario is possible only if the gap between j_1 and j_2 is at least one pixel, it follows that $j_1 < j_2 - 1$. In particular, we obtain that $j_1 < j_2$ and $f_L(i, j_1) < f_L(i, j_2)$.

Claim: Let $p_f(i, j_f)$ and $p_b = (i, j_b)$ be two pixels in the reference view. Pixel p_f occludes p_b in the original image pair *iff* $(i, f_L(i, j_f))$ occludes $(i, f_L(i, j_b))$ after removing the seam pixels.

Proof: We have to show that: $j_b < j_f$ and $j_f + D(i, j_f) = j_b + D(i, j_b)$ *iff* $f_L(i, j_b) < f_L(i, j_f)$ and $f_L(i, j_b) + \hat{D}(i, f_L(i, j_b)) = f_L(i, j_f) + \hat{D}(i, f_L(j_f))$. Using the definition of $\hat{D}$ (see Eq. 12), it follows that:

$$f_L(i, j_b) + \hat{D}(i, f_L(i, j_b)) = \\ f_L(i, j_b) + f_R(i, j_b + D(i, j_b)) - f_L(i, j_b) = \\ f_R(i, j_b + D(i, j_b))$$

$$f_L(i, j_f) + \hat{D}(i, f_L(i, j_f)) = \\ f_L(i, j_f) + f_R(i, j_f + D(i, j_f)) - f_L(i, j_f) = \\ f_R(i, j_f + D(i, j_f)).$$

$$(15)$$

Now, using Lemma 1 we obtain that

$$j_b < j_f \Leftrightarrow f_L(i, j_b) < f_L(i, j_f), \quad \text{and,} \quad (16)$$

$$f_R(i, j_b + D(i, j_b)) = f_R(i, j_f + D(i, j_f)) \Leftrightarrow \\ j_b + d(j_b) = j_f + d(j_f). \quad (17)$$

To complete the proof, the above equations are put together:

$$f_L(i, j_b) + \hat{D}(f_L(i, j_b)) = f_L(i, j_f) + \hat{D}(i, f_L(i, j_f)) \Leftrightarrow \\ j_b + D(i, j_b) = j_f + D(i, j_f).$$

6 EXPERIMENTS & RESULTS

We tested our method on challenging indoor and outdoor scenes. In all experiments, we used the OpenCV implementation of the SGM stereo algorithm [5] to compute the input disparity; hole filling was performed on regions for which the disparity was not computed. This is done by a simple interpolation along scanlines – the holes in each scanline are filled with the minimal disparity value on the hole boundary. The algorithm was implemented in MATLAB and the code as well as the datasets are publicly available.

6.1 Datasets

The following datasets were considered:

Middlebury: Six of the *Middlebury* stereo datasets [19]: *Moebius* (Figure 6), *Aloe* (Figure 7), *Cloth*, *Wood*, *Dolls*, and *Laundry*. These datasets are challenging because the scenes are highly textured and contain objects at different depths. In most of these datasets about 20% of the pixels in the original reference images cannot be removed, since they are either occluding or occluded (see Figure 3).

	V1 & V2	V1 & V5	V1 & V6
Cloth	4.14%	18.06%	23.13%
Wood	5.65%	20.90%	26.61%
Dolls	3.14%	17.70%	23.34%
Laundry	4.44%	20.27%	24.59%

TABLE 1

Initial Occlusions. For each dataset and each baseline, the initial percentage of occluding and occluded pixel out of the total image pixels.

Portrait: A pair of images (Figure 8), provided by [20]. The main challenge in this pair is that the salient object, which covers most of the image, should not be distorted. Moreover, a significant part of the left image is out of the field of view of the right camera, and hence cannot be removed by our algorithm.

Flickr: A set of stereo images, with large depth range, downloaded from *Flickr* (Figure 12-11). The images were manually rectified using [21].

6.2 Geometric Evaluation

A main contribution of our method is the production of a geometrically consistent retargeted image pair that preserves the original depth values of the remaining points. We evaluate depth distortion by measuring the deviation of the updated disparity values from their original values. Our evaluation scheme is described in Figure 1: a disparity map, $\widehat{D}_{SGM}$, is computed on the retargeted pair of stereo images. The computed map, $\widehat{D}_{SGM}$, reflects the geometry that can be recovered from the pair of retargeted images, regardless of the method used to produce them.

The depth distortion is measured by comparing the disparity value of each pixel in $\widehat{D}_{SGM}$ with its original value. In particular, we compute $\widehat{D}_o$, which consists of the original disparity values, D, after removing the relevant seams with respect to the reference view. The absolute difference, $|\widehat{D}_o - \widehat{D}_{SGM}|$, is shown for all our experiments. For comparison, we evaluate the depth distortion caused by independent single image retargeting (see Figure 1**a**). For quantitative evaluation, we define the *depth distortion* score to be the percentage of pixels whose depth, $\widehat{D}_{SGM}$, has been changed by more than one pixel. That is,

$$B = \frac{1}{N} \sum_{(i,j)} \left(|\widehat{D}_o(i,j) - \widehat{D}_{SGM}(i,j)| > 1 \right). \quad (18)$$

Note that the true depth distortion should have been measured directly in 3D rather in 2D (see [22]). However, evaluating the distortion in 3D is not applicable in our case since the cameras are not calibrated and the units of the inverse disparity are unknown. Hence, we choose to compute the disparity distortion, which is correlated with the 3D distortion. Hence, we choose to compute the disparity distortion, which is correlated with the 3D distortion.

6.3 Main Test

We tested our algorithm on the abovementioned datasets using a fixed set of parameters for the 3D weight: $\beta = 0.08$ and
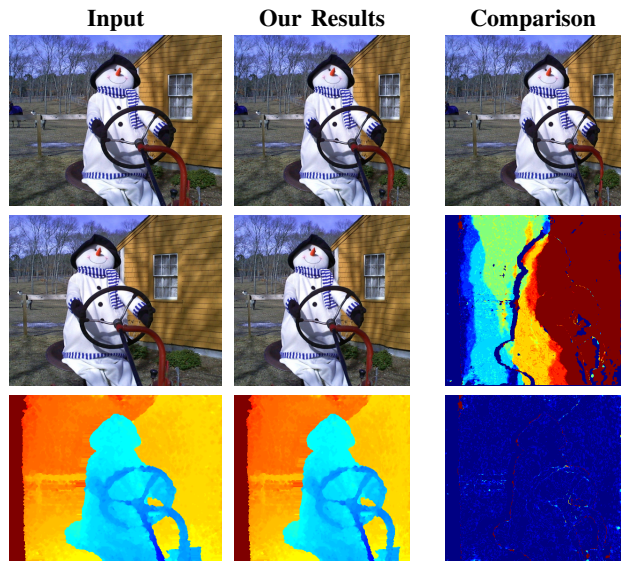

| Input | Our Results | Comparison |

Fig. 10. **Snowman Dataset from Flickr.** See caption of Fig. 6. Depth distortion scores: single image SC, B=70%; stereo pair SC, B=1.3%.

$\gamma = 0.5$ (see Eq. 9). The parameter α (see Eq. 9) was empirically set in the range of 1-5, for each of the datasets. The image width was reduced by 20% for the Middlebury datasets *Aloe* and *Moebius*, and by 17% for the rest. The results are presented in Figures 6-9. In addition, Figures 9-12 show our results, and the results of single image seam-carving, for enlarging the width by 17%.

Our experiments show that the output pair is geometrically consistent and the original depth values are preserved. It is evident that significant depth distortion is caused when naive independent retargeting of each image is considered. (See right columns in each of the figures.)

To evaluate the appearance distortion, we show the single image seam carving result of the left image [4]. The large number of geometric constraints that our method must satisfy limits the number of candidate seams; the constraints are thus expected to yield results that are not as good as those obtained for single image retargeting. Still, the 3D information and the use of generalized seams compensate for this problem. Our results are similar (e.g., Figure 6 and Figure 7) to those of single image seam carving and in some cases much better. For example, our method successfully preserves the face appearance (Figure 8) as well as the face depth (Figure 8**c**), without prior knowledge, such as face location, used by [7]. Figure 12 shows another example in which the perspective of the running track is nicely preserved and the man is not deformed, in contrast to the single image seam carving.

6.4 A Naive Use of The Disparity

A naive approach to using the disparity map is to retarget the reference image (by applying single image seam carving), and map the selected seams to the other image using the disparity map. Figure 2 presents the retargeted images (**d-e**). The ground truth disparity map of the original pair, **c**, is compared to the disparity map computed using the SGM on the naive retargeted pair, **f**, and to our results, **g**. This comparison clearly shows

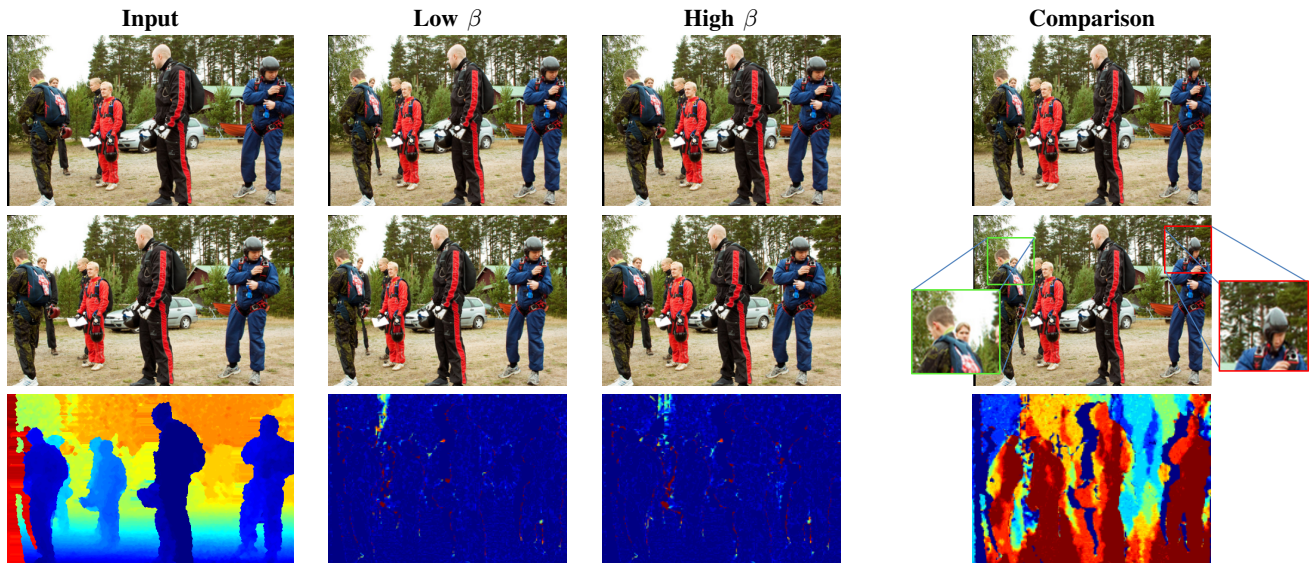| Input	Low β	High β	Comparison

Fig. 11. **People Dataset.** The first column shows the input pair and the computed disparity map. The second and third columns show our results using low and high weights, respectively; the bottom figures show the depth distortion, respectively. The third column shows the results of applying single image SC to the left (top) and right (middle) input images; the bottom figure shows the depth distortion caused by single image SC. Depth distortion scores: single image SC, B=79%; stereo pair SC, low β, B=2.1%; stereo pair SC, high β, B=2.6%.

that the a naive use of the disparity map is insufficient for obtaining geometric consistency. (The retargeted images using our results are presented in Figure 7.)

6.5 3D Weight

So far we have used fixed parameters for the 3D weight. However, as can be seen in the second column of Figure 11, the man on the left almost "lost" his leg. Allowing user interaction for setting the weight of objects according to their depth (the parameter β in Eq. 9) improves the results (third column in Figure 11). The head of the person on the right was not distorted regardless of this parameter, in contrast to the single image seam carving which distorts it dramatically. We note that the geometric consistency of the retargeted images is obtained regardless of the choice of parameters, and depth distortion remains negligible.

Finally, we compare the disparity computed by our method, $\hat{D}$, with that computed by the SGM algorithm on the retargeted images, $\widehat{D}_{SGM}$. For all the tests described above, the difference is less than 2%.

6.6 Effect of Baseline & Retargeting Rate

Next we tested the effect of the baseline between the cameras on the visual appearance and geometric consistency of our results. To this end, we used four of the *Middlebury* datasets: *Cloth*, *Wood*, *Dolls*, and *Laundry*. We tested our method on each dataset, while considering three baselines w.r.t. same view (view1). The initial percentage of occluding and occluded pixels out of the total image pixels is summarized in Table 1. Note that after retargeting, the percentage of occluding and occluded pixels is defined by the new image image size

(since occluding/occluded pixels are not removed). Figure 13 shows the computed *depth distortion* scores (see Eq. 18) for shrinking the image width by 17%, 25%, and 35%. As can be seen, the distortion scores increase to some extent as the baseline is extended, and the image width is reduced. The main reason for this result is that the quality of the disparity map degrades as the baseline is extended. The depth distortion score ranges from 0.039%-3.8%, and the average over all scores (all datasets, baseline, and image widths) is only 0.78%. That is, the depth values of more than 99% of all pixels on average, have been perserved.

In terms of geometric consistency, there is no real restriction regarding the capacity to shrink. An exception is the extreme case where all pixels in a certain row in the left image are occluding/occluded pixels. In such a case (which did not happen in any of our experiments) there is no valid seam to remove. In terms of appearance, the limitation of the capacity to shrink depends also on the occluding and occluded pixels, which are not removed by our method. The number of possible candidate seams is reduced as more seams are removed (the percentage of occluding/occluded pixels is increased); therefore, the visual artifacts are expected to be more significant (see next section).

6.7 Visual Effect of The Input Disparity

As explained above, our method depends on the quality of the disparity map. Errors in the disparity map affect the depth distortion scores as well as the visual appearance of the images. The effect of these errors become more significant as more seams are removed. Figure 14 shows our results on the *Cloth* and *Wood* datasets when shrinking them by 17%, 25%, and 35%. The baseline here is fixed (V1 & V5). In the
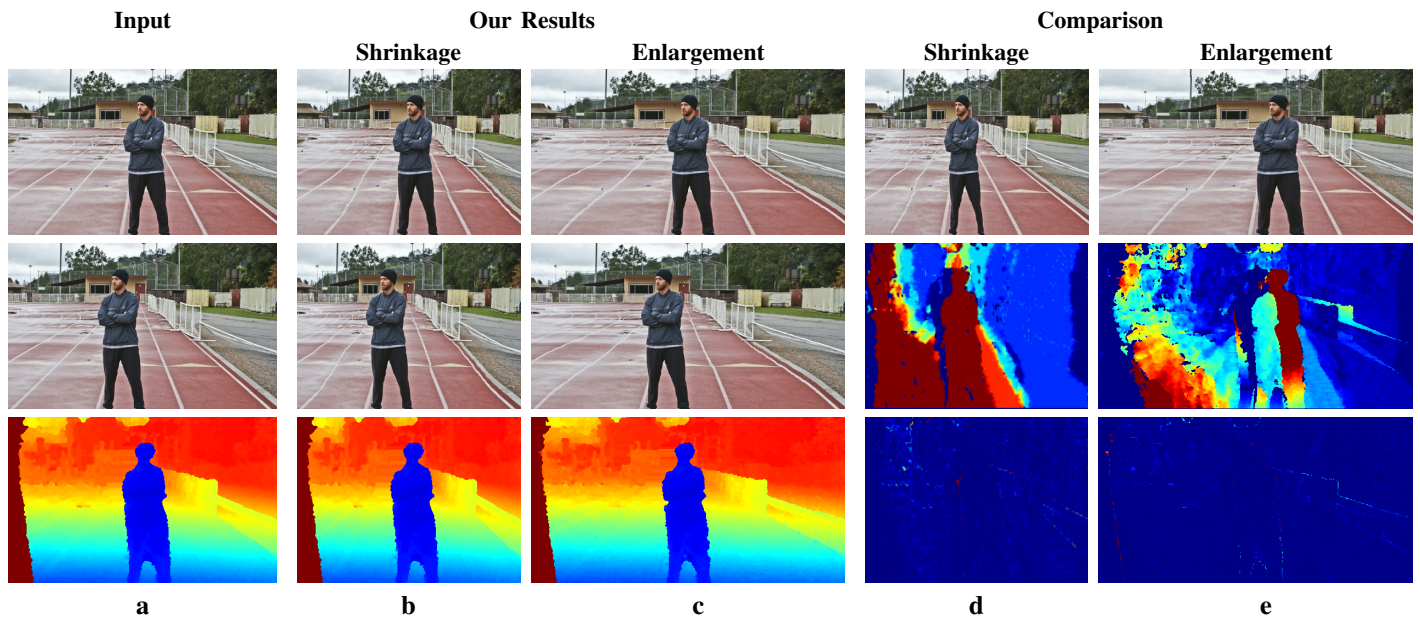
Fig. 12. **Man Dataset.** In column **(a)** (top to bottom), the input left and right images and the input disparity map. In columns **(b)**,**(c)**, our results for reducing and increasing the width in 17%, respectively. In columns **(d)**,**(e)**, the results of applying single image SC to the left input image (top); the distortion in depth caused by independent single image retargeting (middle); the distortion in depth caused by our stereo retargeting method. Depth distortion scores: Shrinkage: single image SC, B=49%; stereo pair SC, B=0.43%; Expansion: single image SC B=62.4%; stereo pair SC, B=1.02%.

left column, the input disparity maps were obtained by the SGM stereo algorithm (followed by hole filling). In the right

Fig. 13. **(a)-(d):** Depth distortion scores for the *Middlebury* datasets: *Cloth*, *Wood*, *Dolls*, and *Laundry*. In each graph, the depth distortion scores (y-axis) computed for three baselines: view1 and view2 (V1 & V2), view1 and view5 (V1 & V5), view1 and view6 (V1 & V6), and for shrinkage of: 17%, 25%, and 35%.

column, the ground truth disparity maps were used as input. The results on the *Cloth* dataset show only a small difference in the appearance when the ground truth disparity was used. Overall, the visual appearance of the images in both cases is preserved. The results on the *Wood* dataset using SGM shows several visual artifacts that becomes more significant as more seams are removed. These distortions do not appear when the ground truth disparity is used.
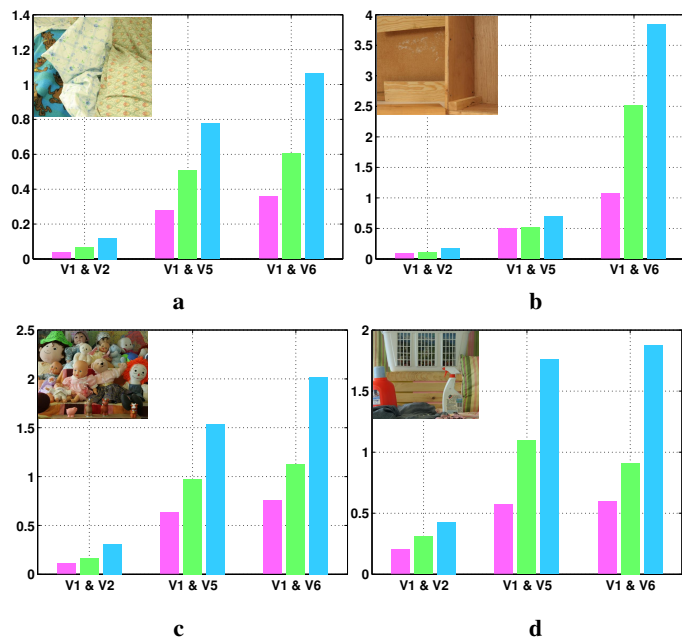
6.8 Comparison with Stereo Retargeting Methods

We compared our method with existing warping-based methods for stereoscopic image retargeting [12], [13]. The comparison of the visual appearance for the same datasets is found in [13]. For all methods, the visual appearance of the left and right images is overall pretty good. In each of the methods, including ours, small visual artifacts can be found. However, we focus here on evaluating the geometric consistency of the results, which is the primary goal and contribution of our method.

Each retargeted stereo image pair gives rise to a new retargeted scene, which is represented by a disparity map between the retargeted images. Figure 17 presents the disparity maps computed by SGM on the input images, and the retargeted pairs obtained by our method, Chang *et al.* [12], and Lee *et al.* [13]. As can be seen, the disparity maps computed on Chang *et al.*'s images (Figure 17**c**) are noisy and contain significant artifacts of both background and foreground objects (e.g., the man on the right in the *People* dataset result). Lee *et al.*'s disparity maps (Figure 17**d**) are less noisy, but the depth values and the 3D structure of the 3D scene are often
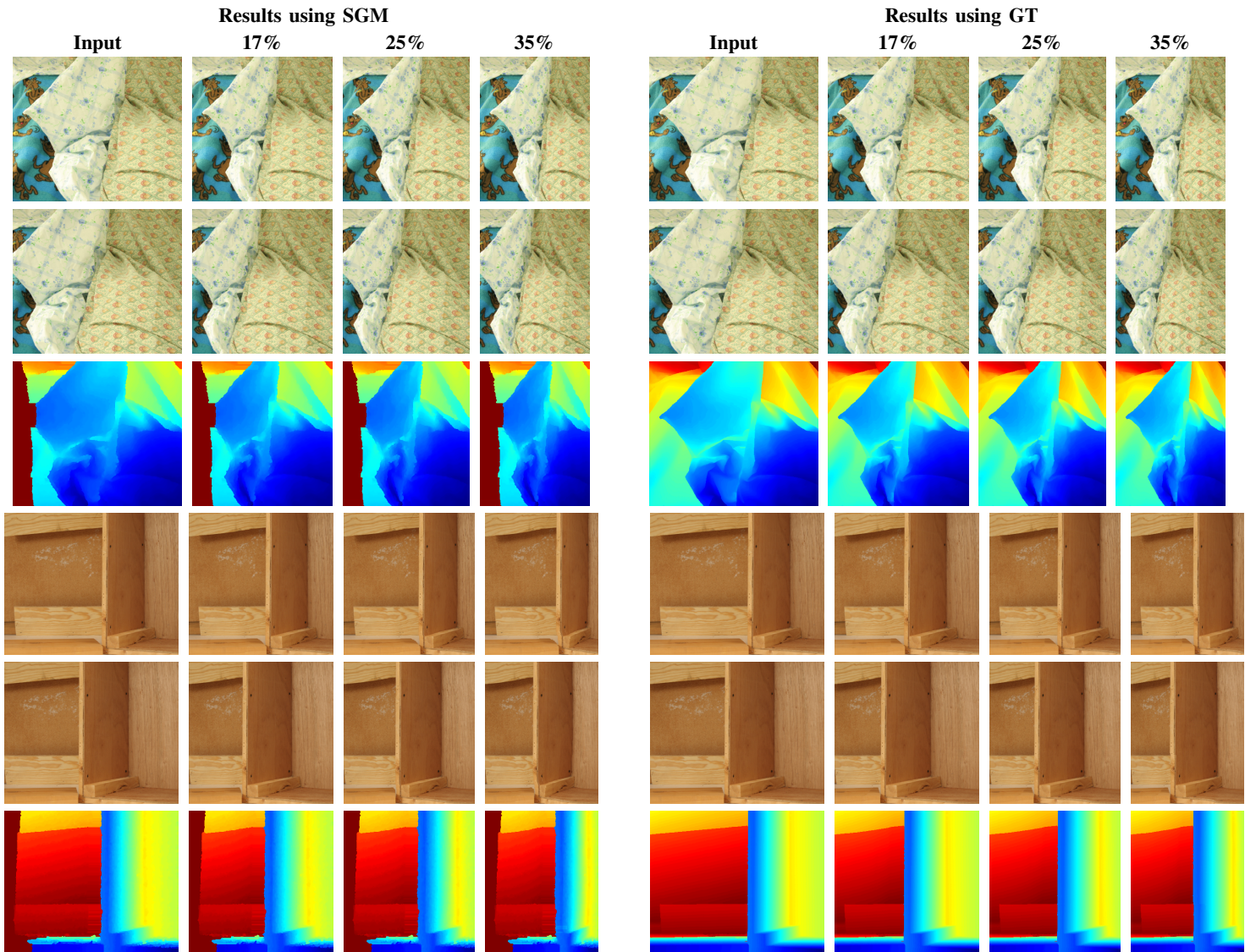
Fig. 14. The results on the *Cloth* and *Wood* datasets for shrinking the image width by: 17%, 25% and 35%. On the left column, the input disparity maps were computed by SGM stereo algoithm. On the right column, the results w.r.t. left column using the ground truth disparity maps as input.

considerably distorted compared to input disparity map. For instance, the change of depth values of the foreground men in the *People* dataset (first row), and the depth change of the face in *Diana* dataset (last row). Another visual artifact can be seen in the results on the *Man* dataset (second row), the 3D structure of running track background (see Figure 12), is significantly bended in Lee's result. None of these artifacts, 3D distortions or noise exist in our results. This comparison demonstrates that neither methods [12], [13] obtained geometrically consistent disparity maps that preserve the original depth values. The depth distortion score for their results cannot be computed since the mapping between the retargeted and input images is not available.

A comparison to Utsugi *et al.* [14] is presented in Figure 7. As can be clearly seen, the appearance of their retargeted images is considerably distorted compared to our results.

6.9 Depth Perception

As demonstrated so far, our method obtains geometrically consistent results, while preserving the original depth values of the 3D structure of the scene. This implies that the depth of the new, retargeted scene should be perceived as similar to the depth of the original scene. To confirm this property, we use anaglyph (red-green) images to perceive the depth of the retargeted 3D scene (see Figure 16).

The perceived depth is often affected by single image 3D cues such as perspective or prior knowledge of the objects' 3D shape. In order to reduce the influence of such cues, we generated a synthetic scene which consists of a background plane and a foreground frame. Both the foreground and the background scene are rendered with the same texture. Figure 15 shows, from left to right, the anaglyph input image, our results for 30% and 50% reduction in the width, and the comparison to the independent single image seam carving. Viewing these images in 3D using red-green glasses
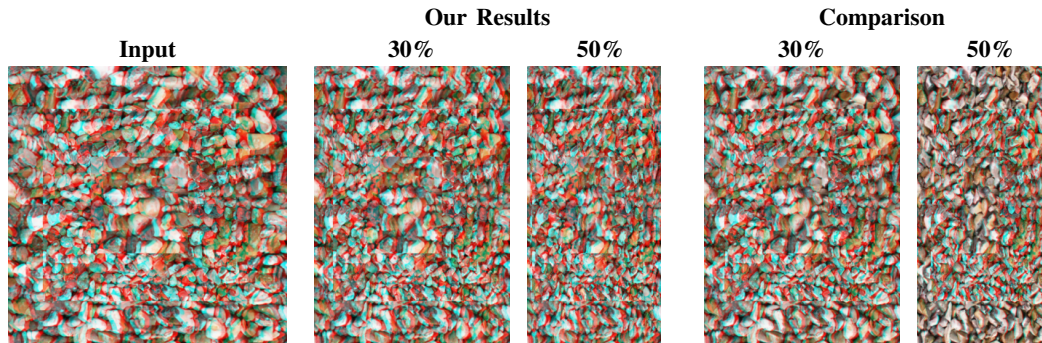
Fig. 15. **Stones Dataset.** Anaglyph images of the input stereo pair, our results for 30% and 50% reduction, and the results of independent single image seam-carving; the images should be viewed with a red-green glasses.

demonstrates that the perceived depth of the retargeted images is similar to the original one (see also Figure 15). Furthermore, when considering the result of independent image retargeting, it is clear that the perceived depth is significantly distorted. The distortion becomes progressively more severe as more seams are removed, until it is impossible to perceive depth. However, it is important to note that our brain can often compensate for the resulting distortions, depending on the 3D structure of scene, the type of objects and other 3D cues.

7 CONCLUSIONS & FUTURE WORK

We extended seam carving to work on a stereo image pair. Retargeting each image independently will distort the geometric structure. We have shown how to extend single image seam carving to work on a pair of images, and proved that the proposed algorithm is guaranteed to give a geometrically consistent result. The retargeted images can thus be viewed on a stereoscopic display or processed by any computer vision algorithm. We demonstrated the advantages of our method on several challenging stereo images, and compared it to current state-of-the-are stereo retargeting methods.

In addition to single image seam carving limitations, our method also affected by the quality of the input disparity map and the amount of occluding and occluded pixels. Both are affected by the texture, the camera locations, and the 3D scene. Another limitation of stereo seam carving is that it can be applied only on rectified stereo pair, and can remove only vertical seams (see Section 4). To reduce the image height, it is necessary to extend the image width and then apply uniform resizing of the image.

On the positive side, in addition to the guaranteed geometric consistency, our method takes advantage of both appearance and depth cues and obtain small appearance distortion in image of scenes that are difficult to deal with using a single image seam carving.

Possible future extensions of our method include extension to stereoscopic video, implementation on smartphones equipped with a stereo camera and 3D display, incorporating depth based saliency map. Another interesting direction is establishing a benchmark for stereoscopic retargeting and editing as well as conducting an in-depth user study for evaluating the depth perception.

REFERENCES

[1] T. Basha, Y. Moses, and S. Avidan, "Geometrically consistent stereo seam carving." in *Proc. Int. Conf. Comp. Vision*, 2011, pp. 1816–1823.
[2] A. Mansfield, P. V. Gehler, L. J. V. Gool, and C. Rother, "Scene carving: Scene consistent image retargeting," in *Proc. European Conf. Comp. Vision*, 2010, pp. 143–156.
[3] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing." in *ACM Trans. Graph.*, 2007, p. 10.
[4] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting." in *ACM Trans. Graph.*, 2008.

| a | b | a | b |

Fig. 16. **Anaglyph images .** (a),(b) the input stereo pair and our results, respectively; the images should be viewed with red-green glasses.

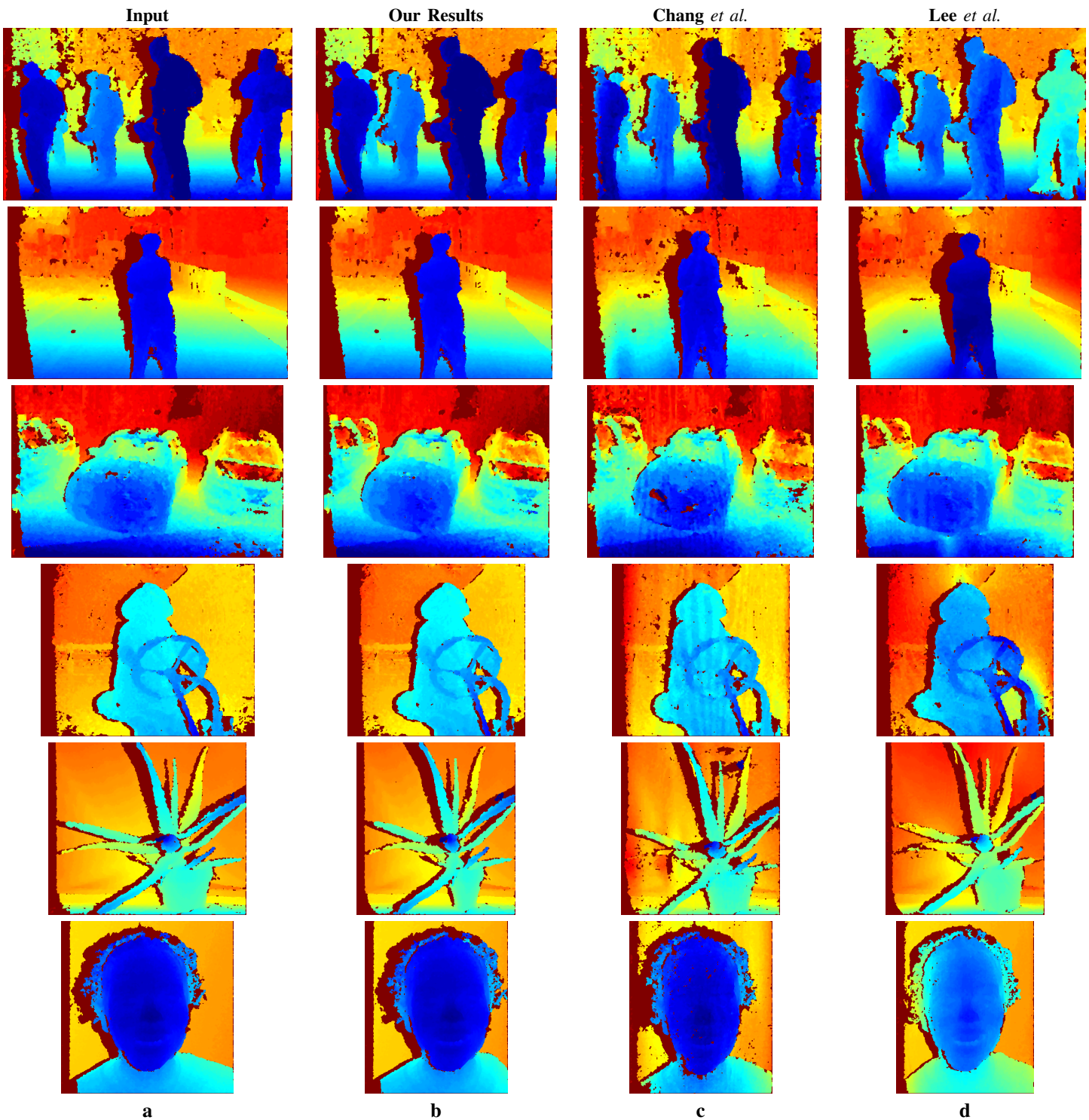| Input	Our Results	Chang *et al.*	Lee *et al.*

a	b	c	d

Fig. 17. **Depth Comparison.** In column (a), the disparity map computed by applying the SGM stereo algorithm on each of the input image pairs: *People*, *Cars*, *Man*, *Snowman*, *Aloe*, and *Diana*. In columns (b-d), the disparity maps computed by applying SGM on the retargeted stereo pairs obtained by our method, Change *et al.*[12], and Lee *et al.*[13], respectively.

[5] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information." in *IEEE Trans. Patt. Anal. Mach. Intell.*, 2008, pp. 328–341.

[6] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. Int. Conf. Comp. Vision*, 2009, pp. 151–158.

[7] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. Int. Conf. Comp. Vision*, 2007, pp. 1–6.

[8] Y.-S. Wang, C.-L. Tai, O. Sorkine, T.-Y. Lee, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing." in *ACM Trans. Graph.*, 2008, p. 118.

[9] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative

study of image retargeting." in *ACM Trans. Graph.*, 2010, p. 160.

[10] M. Grundmann, V. Kwatra, M. Han, and I. A. Essa, "Discontinuous seam-carving for video retargeting." in *Proc. IEEE Conf. Comp. Vision Patt. Recog.*, 2010, pp. 569–576.

[11] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. H. Gross, "Nonlinear disparity mapping for stereoscopic 3d." in *ACM Trans. Graph.*, 2010.

[12] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images." in *Trans. on Multimedia*, 2011, pp. 589–601.

[13] K. Lee, C. Chung, and Y. Chuang, "Scene warping: Layer-based stereoscopic image resizing," in *Proc. IEEE Conf. Comp. Vision Patt. Recog.*, 2012, pp. 49–56.

[14] K. Utsugi, T. Shibahara, T. Koike, K. Takahashi, and T. Naemura, "Seam carving for stereo images," in *3DTV-Conference*, 2010, pp. 1–4.

[15] C. Birklbauer and O. Bimber, "Light-field retargeting," in *Comp. Graph. Forum*, 2012, pp. 295–303.

[16] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images." in *Proc. IEEE Conf. Comp. Vision Patt. Recog.*, 2008.

[17] W.-Y. Lo, J. van Baar, C. Knaus, M. Zwicker, and M. H. Gross, "Stereoscopic 3d copy & paste." in *ACM Trans. Graph.*, 2010, p. 147.

[18] S. J. Koppal, C. L. Zitnick, M. F. Cohen, S. B. Kang, B. Ressler, and A. Colburn, "A viewer-centric editor for 3d movies." in *IEEE Computer Graphics and Applications*, 2011, pp. 20–35.

[19] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comp. Vision Patt. Recog.*, 2007, pp. 1–8.

[20] F. Huguet and F. Devernay, "A variational method for scene flow estimation from stereo sequences," in *Proc. Int. Conf. Comp. Vision*, 2007, pp. 1–7.

[21] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs." in *Mach. Vis. Appl.*, 2000, pp. 16–22.

[22] T. Basha, Y. Moses, and N. Kiryati, "Multi-view scene flow estimation: A view centered variational approach," in *IJCV*, 2012, pp. 1–16.

Tali Basha is currently a PhD student at the school of electrical engineering, Tel-Aviv University, Israel. She received the B.Sc. degree in electrical engineering from Tel-Aviv University, Israel, in 2007, (cum lauda). She then continued to the direct track toward the M.Sc. degree in 2007 and to the direct track toward PhD degree in 2009. She spent the summer of 2010 as an intern in Disney Research, Zurich. Her research interests include image and video analysis using multi-view systems, 3D structure and 3D motion estimation.

Yael Moses received the M.Sc. and Ph.D. degrees in computer science from the Weizmann Institute in 1986 and 1994, respectively. She is currently a senior lecturer in the Efi Arazi School of Computer Science, the Interdisciplinary Center (IDC), Herzliya. She is a founder of IDC's Keren Or program, which makes higher education accessible to students from underprivileged areas. She was a postdoctoral fellow in the Engineering Department at Oxford University, from 1993 to 1994 and at the Weizmann Institute of Science from 1994 to 1998. She spent a sabbatical year at the University of New South Wales and the National ICT Australia, Sydney. Her main research interests are the theory and applications of computer vision. Recently, she has been working on various aspects of multi-camera systems, and distributed computer vision. She is a member of the IEEE

Shai Avidan Shai Avidan received the Ph.D. degree from the School of Computer Science, Hebrew University,Jerusalem, Israel, in 1999. Currently, he is an Assistant Professor at the Faculty of Engineering, Tel-Aviv University, Tel-Aviv, Israel. In between he worked for Adobe, Mitsubishi Electric Research Labs (MERL) and Microsoft Research. He published extensively in the fields of object tracking in video and 3-D object modeling from images. He is also interested in Internet vision applications such as privacy preserving image analysis, distributed algorithms for image analysis, and image retargeting.